The Use of Principal Components Analysis (PCA) in Processing AIRS Data

Mitchell D. Goldberg
Walter Wolf, Lihang Zhou
NOAA/NESDIS/Office of Research and Applications
05/22/2003

Applications of Eigenvectors to AIRS Data

- NESDIS is processing and distributing AIRS data and products in near-real time to NWP centers.
- Data size is very large compared with current sounders (1 orbit ~ 2GB vs. 8 MB)
- Information is not independent. Eigenvectors provide an effective way to extract independent pieces of information.
- Eigenvector expansion coefficients (principal component scores) can be provided instead of the individual channels.
- Individual channels can be reconstructed with minimal signal loss.

NWP Users

- NCEP
- ECMWF
- Met. Office
- Meteo-France
- Goddard DAO
- Meteor. Service of Canada
- Bureau of Meteorology Research Centre (Australia)
- FNMOC

Principal Component Analysis

- Principal component analysis (PCA) is often used to reduce data vectors with many components to a different set of data vectors with much fewer components that still retains most of the variability and information of the original data

- $\mathbf{R} \;=\; r_1.\mathbf{i}_1 + r_2.\mathbf{i}_2 + r_3.\mathbf{i}_3 + \ldots. + r_n.\mathbf{i}_n$

  where $\mathbf{i}_1 = (1,0,0,0,0\ldots.0_n)$; $\mathbf{i}_2 = (0,1,0,0,0\ldots.0_n)$

- Data are rotated onto a new set of axes, such that the first few axes have the most explained variance.

- $\mathbf{R} \;=\; p_1.\mathbf{E}_1 + p_2.\mathbf{E}_2 + p_3.\mathbf{E}_3 + \ldots. + p_n.\mathbf{E}_n$

  where $\mathbf{E}$ are eigenvectors and $p_1 = \mathbf{R}\,\mathbf{E}_1$

- So instead of **R** vectors of length n, we can have a truncated **P** vectors of length m, where m << n

Eigenvectors are used for

- Data compression
- Quality control
- Retrievals of geophysical parameters
  - » Atmospheric temperature, moisture and ozone
  - » Surface temperature and emissivity

Generating AIRS eigenvectors

- **Each AIRS data vector has 1524 radiance values (Efforts are being made to expand it to over 2000).**
- **The radiances are normalized by expected instrumental noise (signal to noise)**
- **Compute the covariance matrix S**
- **Compute the eigenvectors E and eigenvalues Λ**

$$S = E\,\Lambda E^T$$

- **E = matrix of orthonormal eigenvectors (1524x1524)**

Λ = vector of eigenvalues (explained variance)

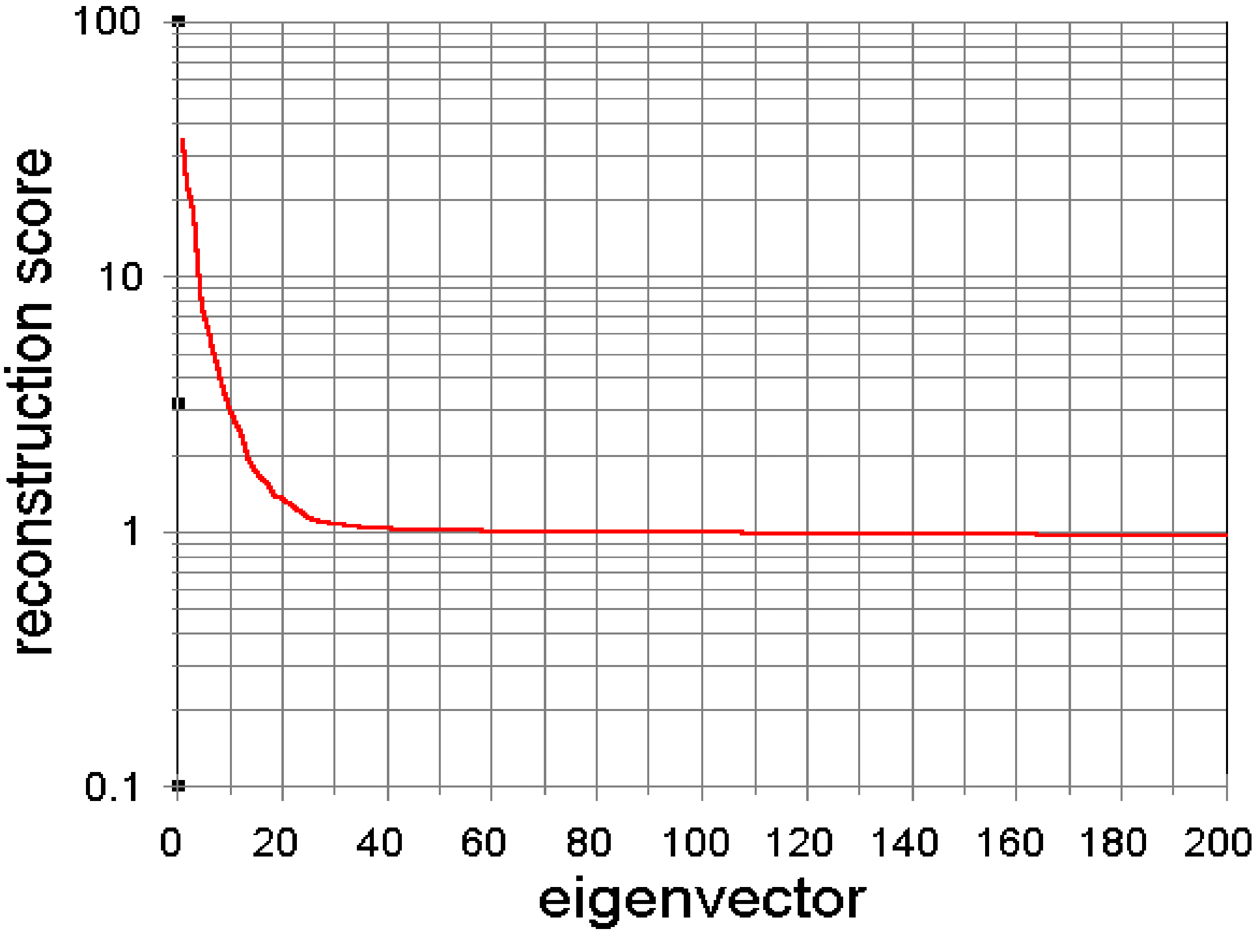
Applying AIRS eigenvectors

- On independent data – compute principal component scores.
- $P \quad = \quad E^T \; R$; elements of $R = (r_i - \bar{r}_i)/n_i$
- Invert equation and compute reconstructed radiances R^*.
- $R^* \quad = \quad E\,P$
- Reconstructed radiances are used for quality control.
- Reconstruction score $= \left[\, 1/N \; 3(R^*_i - R_i)^2 \,\right]^{1/2}$
  $i = 1 \;\ldots.N$ channels

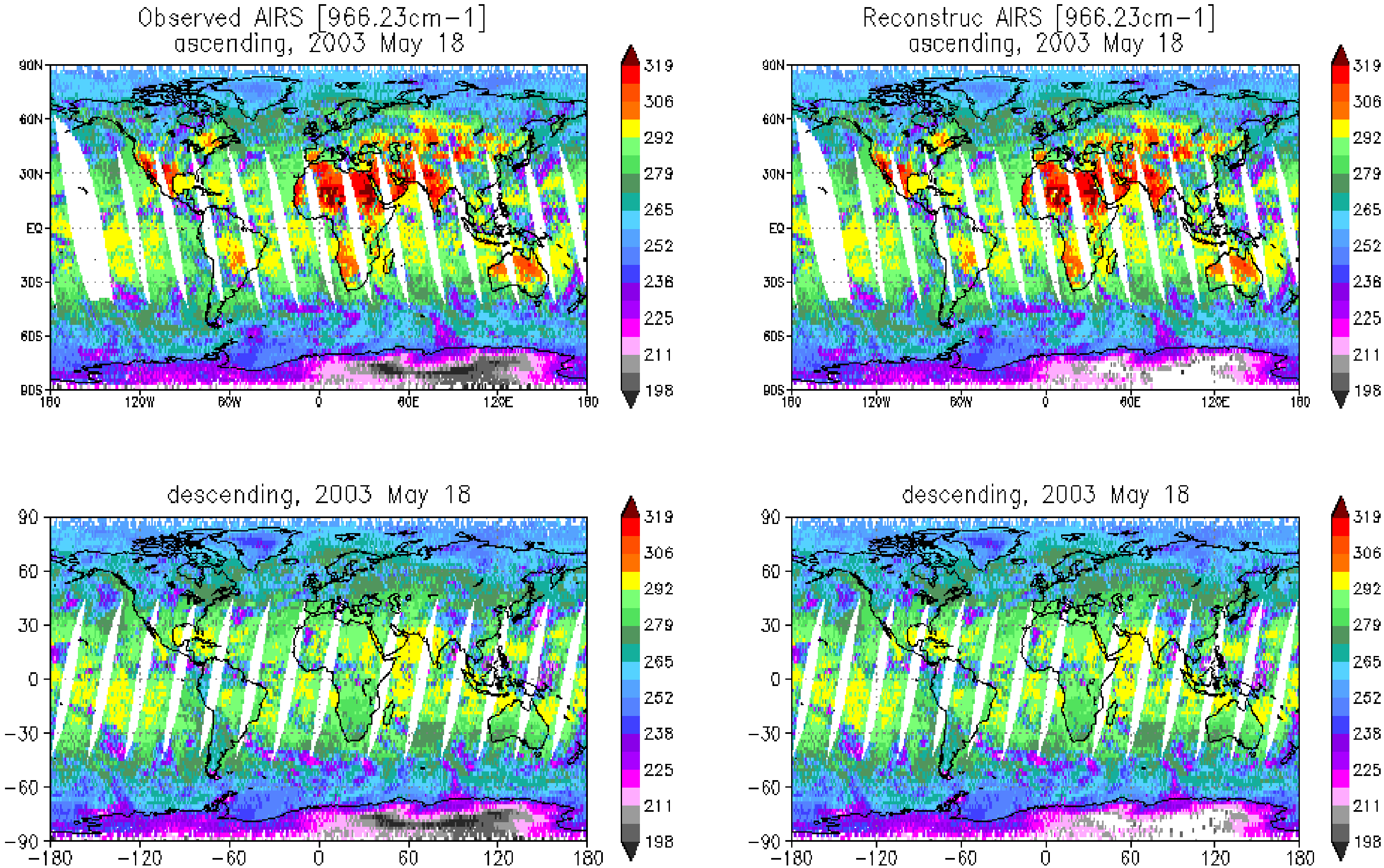
AIRS applications

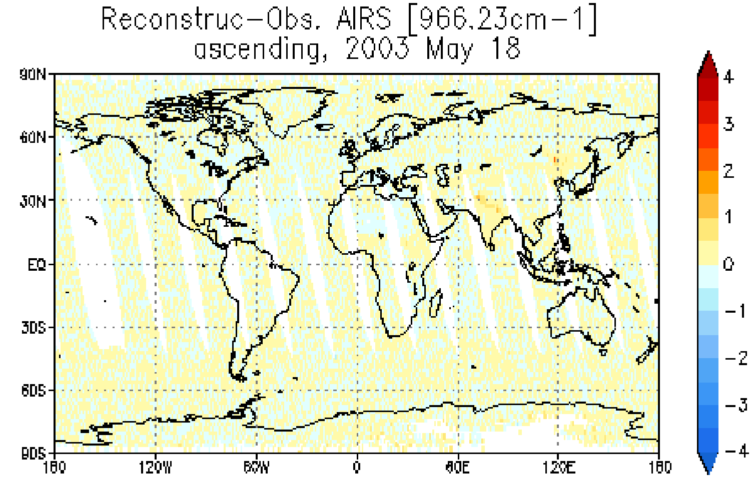
- AIRS is currently being processed and distributed in near-real time.
- Each day a spatial subset of AIRS data is produced (200 mbytes vs 30 GB full data)
- Eigenvectors are generated daily.
- A static set of eigenvectors is used, but the ensemble is occasionally updated with new structures.
- When the ensemble is updated a new set of eigenvectors is also updated.
- Occasional bad channels can be synthesized from neighboring/most correlated channels.

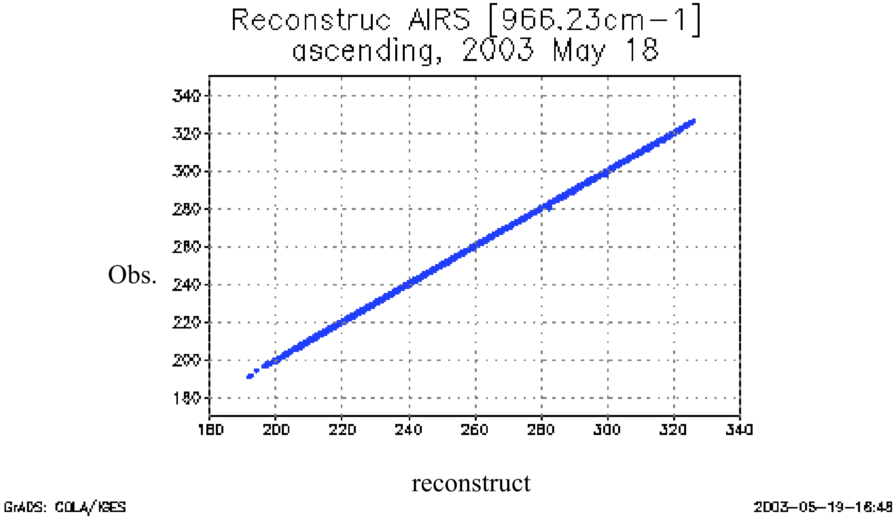
•Reconstruction score = $[\ 1/N \sum(R^*_i - R_i)^2\ ]^{1/2}$
$i = 1 \ldots N$ channels

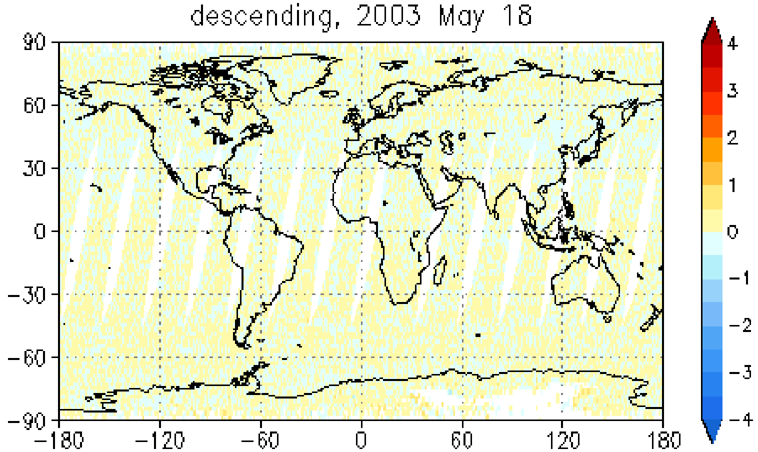


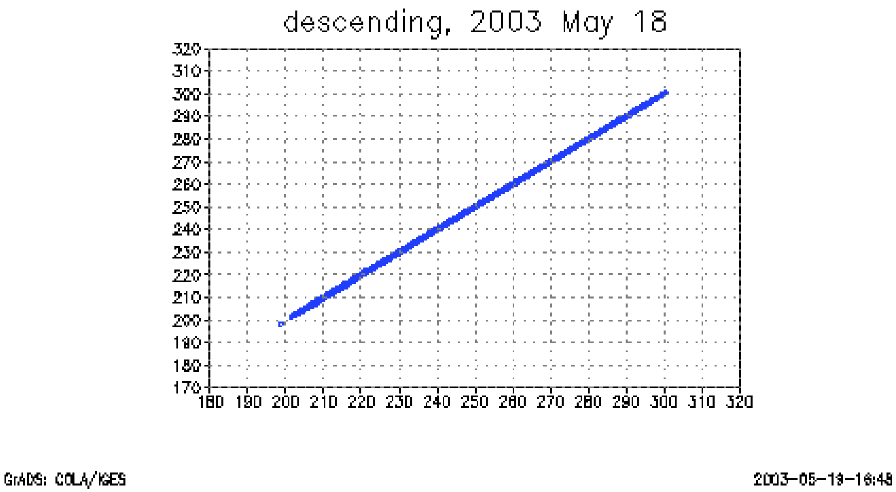
Compare the observed and reconstructed – a window channel




Reconstruc-Obs. AIRS [966.23cm-1]
ascending, 2003 May 18
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180
4
3
2
1
0
-1
-2
-3
-4


Reconstruc AIRS [966.23cm-1]
ascending, 2003 May 18
Obs.
reconstruct
180
200
220
240
260
280
300
320
340
GrADS: COLA/IGES
2003-05-19-16:48

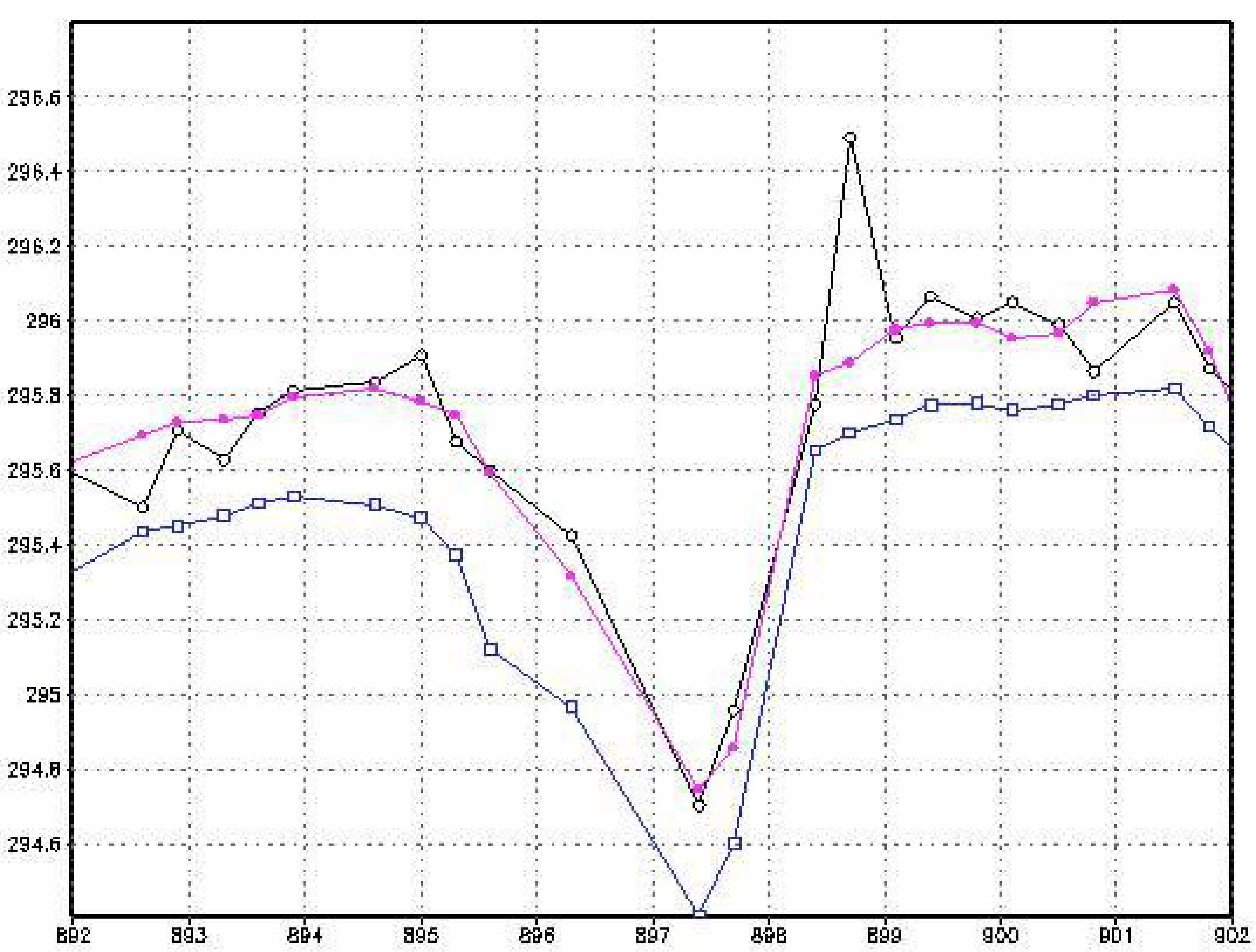

descending, 2003 May 18
90
60
30
0
-30
-60
-90
-180
-120
-60
0
60
120
180
4
3
2
1
0
-1
-2
-3
-4


descending, 2003 May 18
170
180
190
200
210
220
230
240
250
260
270
280
290
300
310
320
GrADS: COLA/IGES
2003-05-19-16:48

Noise filter

- Truncating eigenvectors also results in noise filtering.

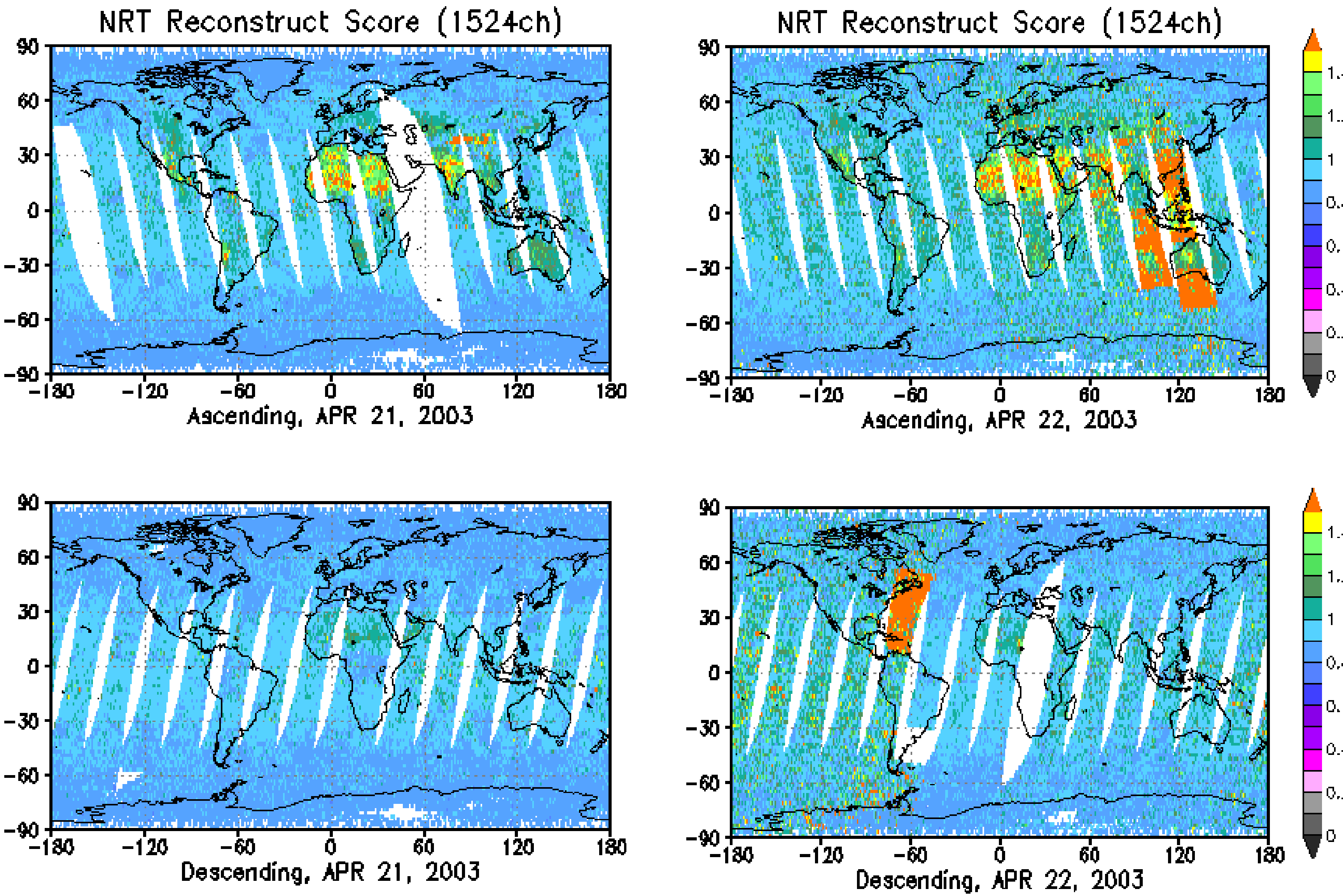
Spectrum from 892 to 902 wavenumber



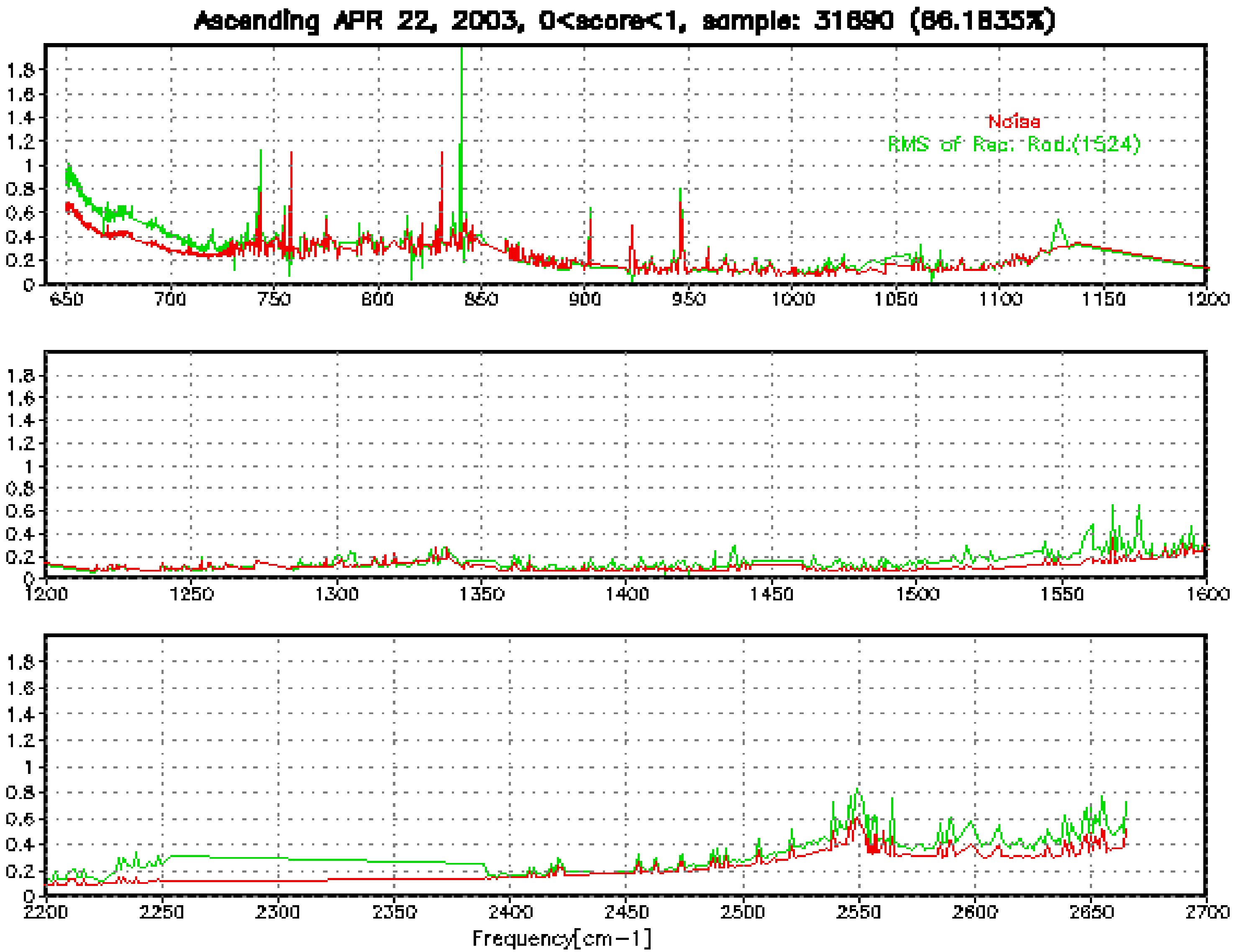
Purple is reconstructed, black is original, blue is calculated from model

Monitoring Eigenvectors

- Monitoring eigenvectors is critical
- Bad channel can be found out by monitoring the reconstruction scores and the difference between the reconstructed and the observed radiances.

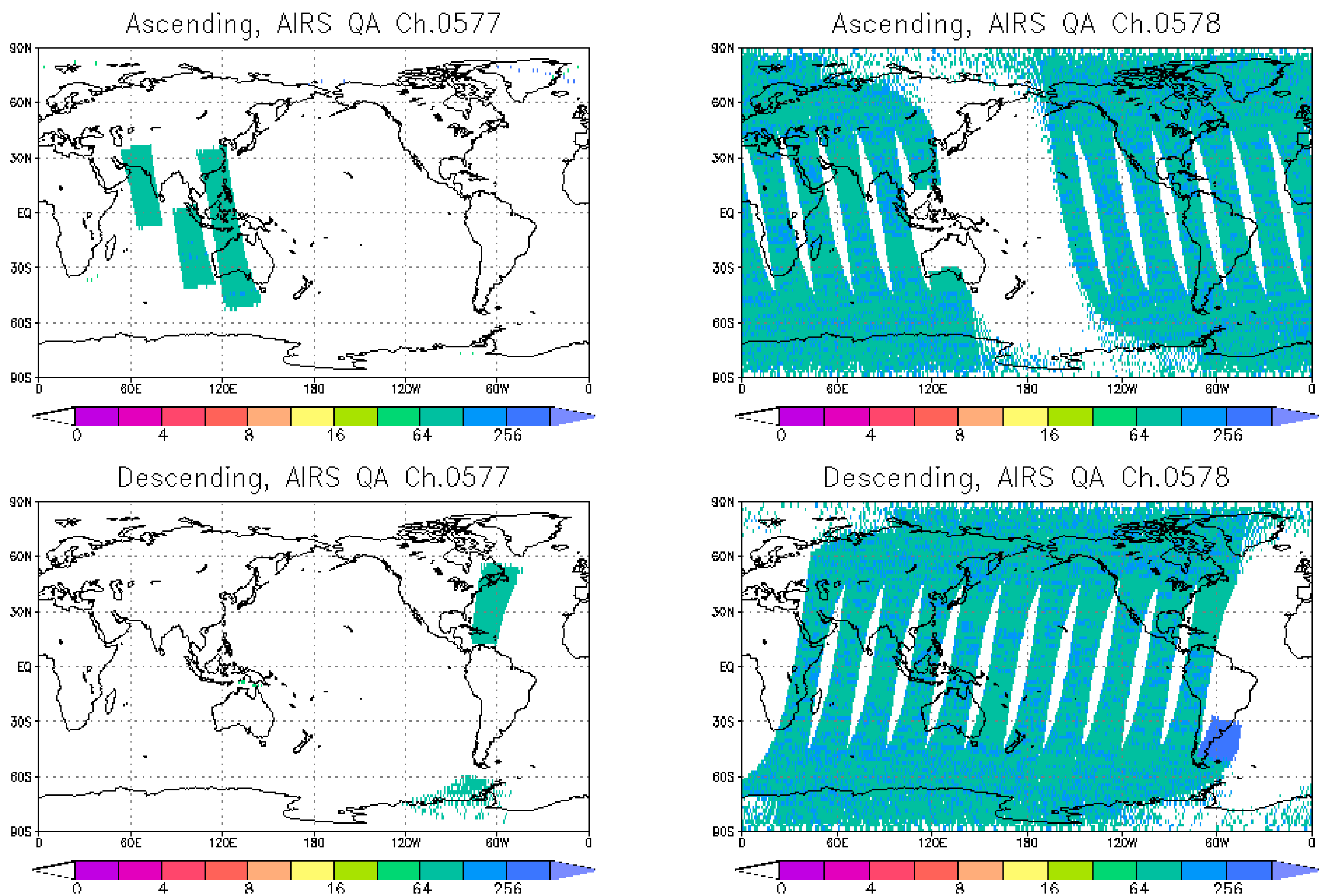

NRT Reconstruct Score (1524ch)
Ascending, APR 21, 2003
NRT Reconstruct Score (1524ch)
Ascending, APR 22, 2003
Descending, APR 21, 2003
Descending, APR 22, 2003
90
60
30
0
−30
−60
−90
−180
−120
−60
0
60
120
180
1.4
1.2
1
0.8
0.6
0.4
0.2
0

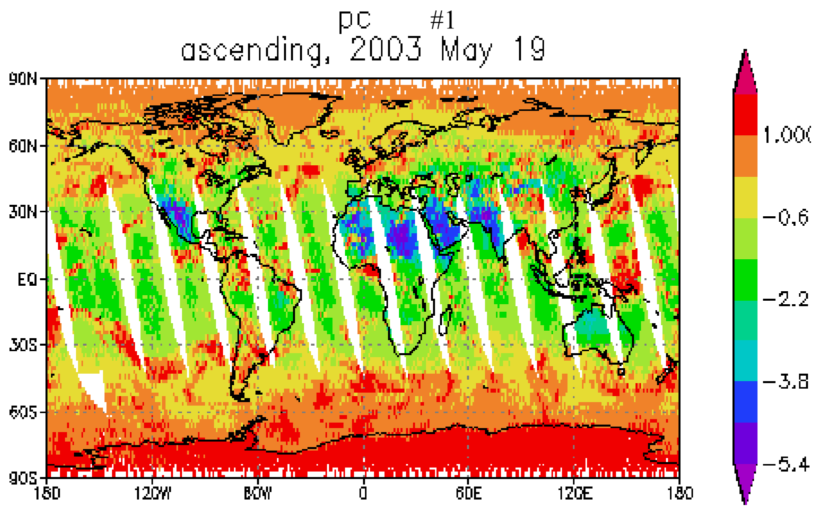
AIRS NEDT vs. RMS of Reconstruct Radiance

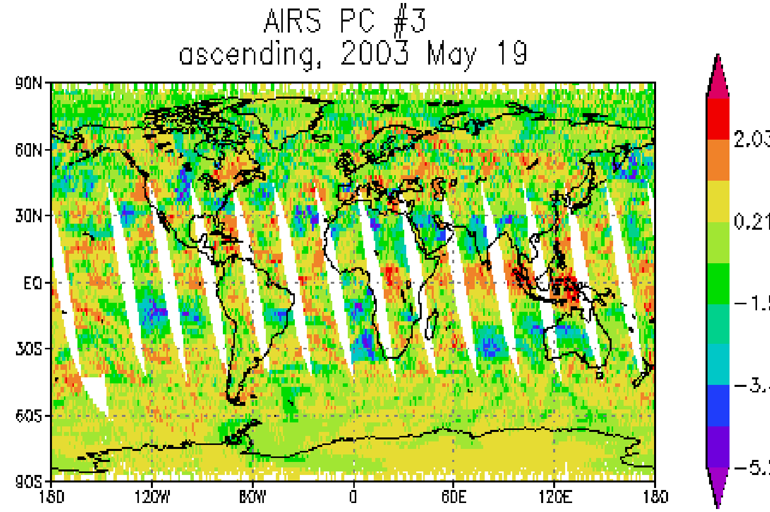


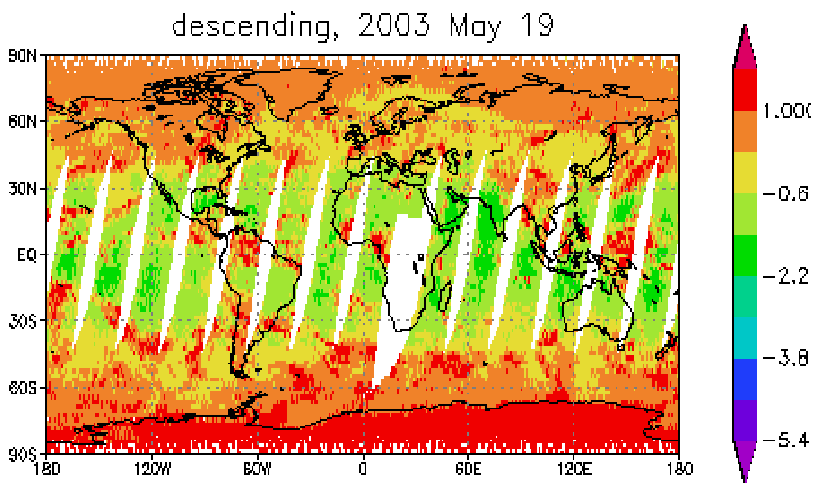
Indx	freq.	Nedt	aveob/avere	s<1 asc	s<1 desc	1<s<1.5 a	1<s<1.5 d	1.5<s<2 a	1.5 <s<2 d	s>2 a	s>2 d
418	832.810	0.356	277.15/ 277.19	0.336	0.336	0.311	0.323	0.350	0.342	0.425	0.447
419	833.201	0.285	277.15/ 277.18	0.300	0.300	0.314	0.318	0.375	0.371	0.385	0.398
420	833.986	0.349	277.16/ 277.20	0.330	0.328	0.298	0.311	0.321	0.333	0.414	0.433
421	835.166	0.314	276.72/ 276.75	0.318	0.321	0.294	0.309	0.325	0.356	0.508	0.507
422	835.560	0.338	276.26/ 276.32	0.376	0.382	0.396	0.396	0.487	0.508	0.599	0.616
423	836.743	0.469	277.13/ 277.15	0.585	0.580	0.658	0.628	0.814	0.796	0.743	0.788
424	837.930	0.281	277.05/ 277.09	0.277	0.278	0.255	0.262	0.278	0.281	0.403	0.428
425	839.121	0.426	276.84/ 276.86	0.437	0.441	0.397	0.405	0.391	0.422	0.465	0.455
426	839.916	0.296	271.72/ 272.47	0.344	0.315	2.346	0.608	9.860	4.562	15.296	5.237
427	840.314	0.305	275.40/ 274.21	2.204	2.311	5.578	6.470	10.692	14.757	22.821	24.518
428	840.712	0.393	276.33/ 276.37	0.379	0.379	0.351	0.375	0.393	0.440	0.605	0.643
429	841.111	0.482	276.76/ 276.83	0.195	0.198	0.226	0.212	0.325	0.262	0.428	0.435
430	842.710	0.542	276.84/ 276.90	0.590	0.589	0.595	0.590	0.683	0.641	0.691	0.804
431	843.913	0.295	276.87/ 276.89	0.289	0.291	0.267	0.276	0.312	0.309	0.408	0.423
432	845.522	0.498	276.64/ 276.68	0.493	0.495	0.453	0.455	0.482	0.508	0.622	0.627
433	845.925	0.287	276.59/ 276.59	0.292	0.292	0.264	0.274	0.296	0.309	0.395	0.358
434	846.733	0.418	276.65/ 276.70	0.412	0.412	0.378	0.386	0.405	0.399	0.452	0.509

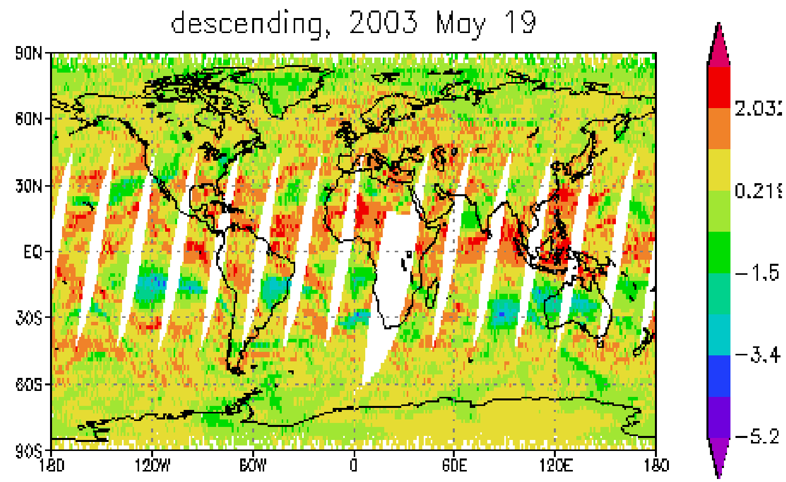
QA Flag: 64- Bad Telemetry 128- High Noise

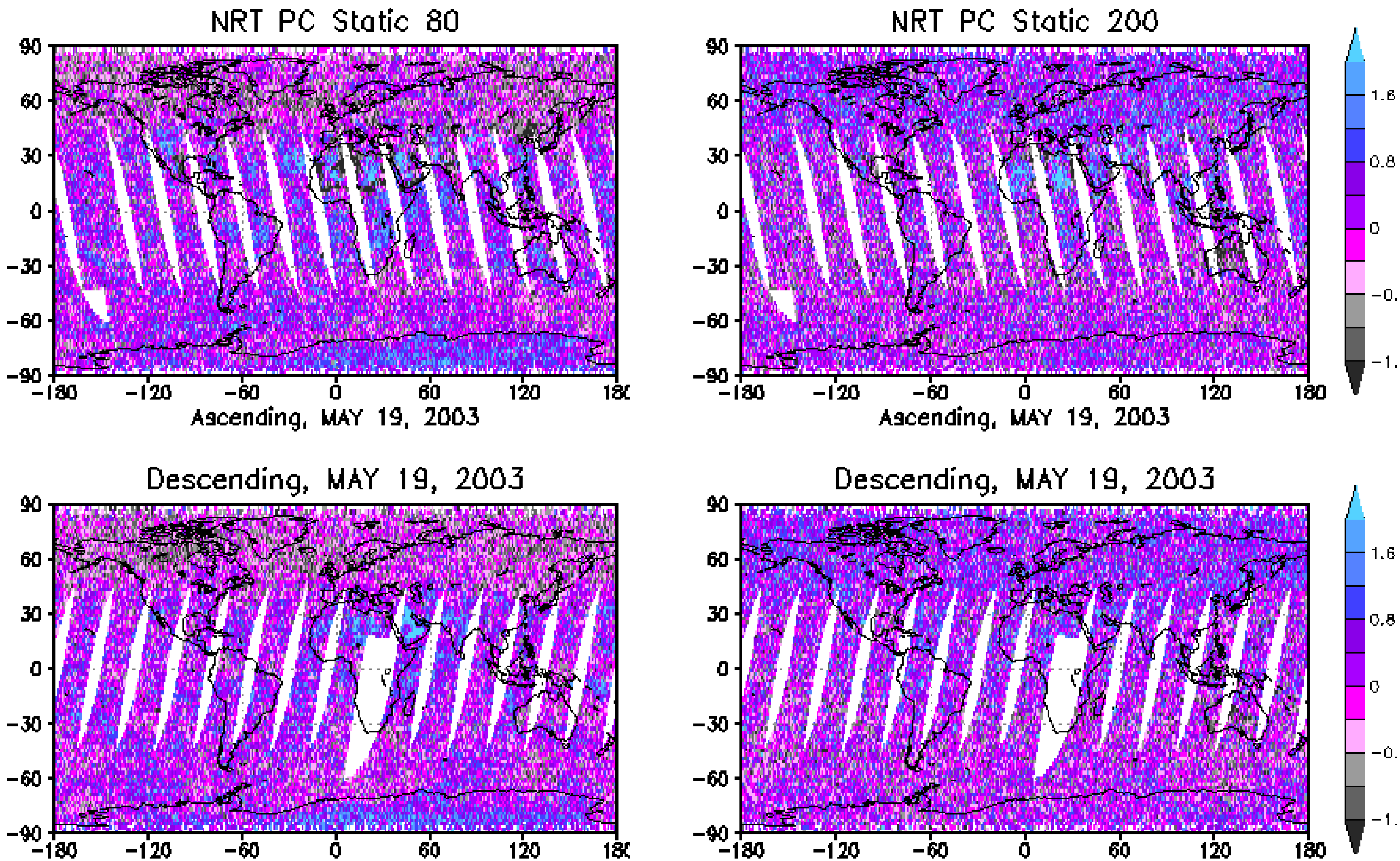



pc #1
ascending, 2003 May 19
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180
1.00
-0.6
-2.2
-3.8
-5.4


AIRS PC #3
ascending, 2003 May 19
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180
2.03
0.21
-1.5
-3.4
-5.2


descending, 2003 May 19
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180
1.00
-0.6
-2.2
-3.8
-5.4


descending, 2003 May 19
90N
60N
30N
EQ
30S
60S
90S
180
120W
60W
0
60E
120E
180
2.03
0.21
-1.5
-3.4
-5.2


NRT PC Static 80
NRT PC Static 200
90 60 30 0 −30 −60 −90
−180 −120 −60 0 60 120 180
Ascending, MAY 19, 2003
Ascending, MAY 19, 2003
Descending, MAY 19, 2003
Descending, MAY 19, 2003
1.6
0.8
0

PC Regression retrievals

1	0.027	240.235	-0.046	2.873	1.196	2840
2	0.057	240.235	-0.046	2.873	1.196	2840
3	0.107	240.235	-0.046	2.873	1.196	2840
4	0.181	243.998	-0.051	2.322	0.952	2840
5	0.365	254.193	-0.010	1.983	0.780	2840
6	0.740	262.260	-0.033	2.286	0.872	2840
7	1.331	259.519	-0.203	2.112	0.814	2840
8	2.513	248.332	-0.010	1.778	0.716	2840
9	5.148	234.652	0.089	1.679	0.716	2840
10	10.707	226.589	-0.065	1.096	0.484	2840
11	21.788	220.779	-0.186	0.941	0.426	2840
12	38.154	215.030	-0.020	0.840	0.391	2840
13	62.214	209.446	0.044	0.828	0.395	2840
14	97.508	206.399	0.033	0.928	0.449	2840
15	154.049	212.950	-0.043	0.791	0.372	2840
16	245.160	224.303	-0.071	0.842	0.375	2840
17	321.809	235.806	0.050	0.878	0.372	2840
18	367.255	242.400	0.023	0.833	0.344	2840
19	425.302	249.901	0.002	0.857	0.343	2840
20	497.472	257.812	0.021	0.833	0.323	2840
21	576.371	264.907	-0.068	0.889	0.336	2840
22	662.038	271.346	-0.173	0.936	0.345	2840
23	754.468	277.064	-0.191	0.973	0.351	2840
24	853.619	281.082	0.006	1.060	0.377	2840
25	959.407	285.599	-0.035	1.090	0.381	2840
26	122.701	0.002	0.000	0.000	1.284	2840
27	166.353	0.000	0.000	0.000	37.853	2840
28	225.145	0.002	0.000	0.001	36.058	2840
29	301.794	0.009	0.000	0.003	36.657	2840
30	401.665	0.034	0.000	0.013	39.428	2840
31	538.612	0.128	0.004	0.049	37.873	2840
32	709.941	0.397	0.017	0.141	35.387	2840
33	894.219	1.293	-0.008	0.188	14.541	2803
34	986.067	1.849	0.013	0.255	13.795	2840
35	1.564	6.538	-0.040	0.158	2.420	2840
36	3.115	11.039	-0.072	0.245	2.218	2840
37	6.121	25.737	-0.242	0.760	2.954	2840
38	11.311	36.662	-0.367	1.227	3.346	2840
39	21.788	61.484	-0.363	2.062	3.354	2840
40	45.055	56.720	-0.206	4.807	8.476	2840
41	97.417	29.089	-0.436	4.581	15.747	2840
42	184.540	15.227	-0.079	3.271	21.480	2840
43	544.011	33.658	-0.420	2.806	8.337	2840
44	852.788	276.156	-2.225	8.628	3.124	2840

45 -2.2216355E-02 0.7892480 2840

1	0.027	239.971	0.306	3.133	1.306	2772
2	0.057	239.971	0.306	3.133	1.306	2772
3	0.107	239.971	0.306	3.133	1.306	2772
4	0.181	243.822	0.264	2.601	1.067	2772
5	0.365	254.300	0.105	2.043	0.803	2772
6	0.740	262.399	0.016	2.072	0.789	2772
7	1.331	259.369	0.169	2.020	0.779	2772
8	2.513	248.846	-0.043	1.847	0.742	2772
9	5.148	235.258	0.033	1.724	0.733	2772
10	10.707	226.898	0.151	1.084	0.478	2772
11	21.788	220.889	-0.257	0.994	0.450	2772
12	38.154	214.895	-0.063	0.748	0.348	2772
13	62.214	209.019	0.035	0.803	0.384	2772
14	97.508	205.800	-0.114	0.907	0.441	2772
15	154.049	212.609	-0.164	0.884	0.416	2772
16	245.160	224.253	-0.072	0.974	0.434	2772
17	321.809	235.963	0.029	0.983	0.417	2772
18	367.255	242.517	0.141	0.860	0.355	2772
19	425.302	250.029	0.185	0.918	0.367	2772
20	497.472	258.070	0.105	0.894	0.347	2772
21	576.371	265.107	0.037	0.807	0.305	2772
22	662.038	271.455	0.018	0.945	0.348	2772
23	754.468	277.265	-0.064	1.032	0.372	2772
24	853.619	281.495	-0.053	1.056	0.375	2772
25	959.407	285.987	-0.097	1.247	0.436	2772
26	122.701	0.002	0.000	0.000	1.572	2772
27	166.353	0.000	0.000	0.000	41.958	2772
28	225.145	0.002	0.000	0.001	36.143	2772
29	301.794	0.009	0.000	0.003	38.434	2772
30	401.665	0.034	0.001	0.014	41.828	2772
31	538.612	0.147	0.002	0.053	35.692	2772
32	709.941	0.459	0.017	0.144	31.425	2772
33	894.219	1.355	-0.004	0.206	15.206	2751
34	986.067	1.998	0.016	0.275	13.781	2772
35	1.564	6.456	-0.017	0.153	2.362	2772
36	3.115	10.953	-0.081	0.269	2.460	2772
37	6.121	25.923	-0.363	0.866	3.342	2772
38	11.311	37.331	-0.532	1.375	3.683	2772
39	21.788	62.395	-0.773	2.489	3.989	2772
40	45.055	56.402	-0.749	5.251	9.310	2772
41	97.417	29.444	-0.293	4.546	15.440	2772
42	184.540	16.019	-0.130	3.427	21.396	2772
43	544.011	33.868	-0.549	3.024	8.929	2772
44	852.788	278.792	-3.487	10.193	3.656	2772

45 -1.0573603E-02 0.7474658 2772

Compression Factors

- 200 pc scores ~ 9.5 % of the total (lossy compression).
- For near lossless compression. Store 80 pc scores plus the residual. The residual can be stored as 1 byte as long as residual is +- 1.28 K. Allows for both noise filtering and lossless compression.
- Not all channels will be within +- 1.28 K, but no more than 20 channels per fov.
- Full data set 2100 x 4 bytes = 9400
- Compression 80 pc x 4 bytes + 2100 ch x 1 byte + 20 outliers x 4 = 2500 ~ 27 %
- Bit trimming can provide even greater lossless data compression (perhaps 10%).

Summary

- All sky eigenvectors
- Monitor reconstruction scores
- Quality control and bad channel handelling
- Provide to NWP centers 200 pc scores (normalized by sqrt(eigenvalues)).
- Also provide subset of individual channels (to monitor the reconstruction).
- Need additional work to study noise filtering feature of eofs.